

A universal expectation bound on empirical projections of deformed random matrices

Kamil Jurczak*

Ruhr-Universität Bochum

October 18, 2012

Abstract

Let C be a real-valued $M \times M$ matrix with singular values $\lambda_1 \geq \dots \geq \lambda_M$ and E a random matrix of centered i.i.d. entries with finite fourth moment. In this paper we give a universal upper bound on the expectation of $\|\hat{\pi}_r X\|_{S_2}^2 - \|\pi_r X\|_{S_2}^2$, where $X := C + E$ and $\hat{\pi}_r$ (resp. π_r) is a rank- r projection maximizing the Hilbert-Schmidt norm $\|\tilde{\pi}_r X\|_{S_2}$ (resp. $\|\tilde{\pi}_r C\|_{S_2}$) over the set $\mathcal{S}_{M,r}$ of all orthogonal rank- r projections. This result is a generalization of a theorem for Gaussian matrices due to Rohde (2012). Our approach differs substantially from the techniques of the mentioned article. We analyze $\|\hat{\pi}_r X\|_{S_2}^2 - \|\pi_r X\|_{S_2}^2$ from a rather deterministic point of view by an upper bound on $\|\hat{\pi}_r X\|_{S_2}^2 - \|\pi_r X\|_{S_2}^2$, whose randomness is totally determined by the largest singular value of E .

1 Introduction

Let C be a real-valued $M \times M$ matrix, $M \in \mathbb{N}$, with singular values $\lambda_k = \lambda_k(C)$, $k = 1, \dots, M$, in decreasing order and E a $M \times M$ random matrix, whose entries are centered i.i.d. real-valued random variables. We denote the singular values of E by $\sigma_1 \geq \dots \geq \sigma_M$. Further let π_r be a rank- r projection, which maximizes the Hilbert-Schmidt norm $\|\tilde{\pi}_r C\|_{S_2}$ over the set $\mathcal{S}_{M,r}$ of all orthogonal rank- r projections into subspaces of \mathbb{R}^M .

Consider the process $(Z_{\tilde{\pi}_r})_{\tilde{\pi}_r \in \mathcal{S}_{M,r}}$ defined by

$$Z_{\tilde{\pi}_r} := \|\tilde{\pi}_r X\|_{S_2}^2 - \|\pi_r X\|_{S_2}^2, \quad X := C + E, \quad (1.1)$$

and its supremum denoted by

$$Z_{\hat{\pi}_r} = \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} Z_{\tilde{\pi}_r}, \quad (1.2)$$

where $\hat{\pi}_r$ is a location of the supremum. In general, $\hat{\pi}_r$ is not unique, since the distribution of the entries is allowed to have mass points.

Rohde (2012) investigates the accuracy for empirical reduced-rank projection of the deterministic matrix C perturbed by a Gaussian noise matrix E . More precisely, she derives

*Supported by the DFG research unit FG 1735, RO 3766/3-1

upper and lower bounds on $\mathbb{E}Z_{\hat{\pi}_r}$. The proofs in the mentioned article rely heavily on the Gaussian distribution of E . In particular, the main ingredients for the upper bound are among others S_2 - S_∞ -chaining and the Borell (1975) - Sudakov and Tsirel'son (1974) inequality. Since Z is not centered, the clue of the paper is a slicing argument for $\mathcal{S}_{M,r}$ to proceed to centered Gaussian processes on well-chosen slices. Beyond, for the proofs of lower bounds on $\mathbb{E}Z_{\hat{\pi}_r}$, the invariance property of the distribution of E under orthogonal transformation and Sudakov's minoration are used. Due to the dependence of the proofs on the Gaussian distribution, naturally the question arises, whether the results of Rohde (2012) hold for a larger class of probability distributions of the independent entries E_{ij} . Before we pursue this question, we first recapitulate the upper and lower bounds from Rohde (2012).

Theorem 1 (Upper bound for Gaussian matrices)

Under the former assumptions and notations let the distribution of E_{ij} be centered Gaussian with variance σ^2 and $\text{rank}(C) \geq r$. Then in case of $r \leq M - r$ the following bound holds

$$\mathbb{E}Z_{\hat{\pi}_r} \lesssim \sigma^2 r M \left(\min \left(\frac{\lambda_1^2}{\lambda_r^2}, 1 + \frac{\lambda_1}{\sigma \sqrt{M}} \right) + \min \left(\left(\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2} \right)^{\frac{1}{2}} \cdot \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \right) \right), \quad (1.3)$$

where $\frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}$ is set to infinity, if $\lambda_r = \lambda_{r+1}$.

Theorem 2 (Lower bounds for Gaussian matrices)

Let E_{ij} be centered Gaussian with variance σ^2 .

(i) Let $\lambda_1 = \dots = \lambda_M = \alpha$, then

$$\mathbb{E}Z_{\hat{\pi}_r} \geq \mathbb{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2 \right) \quad (1.4)$$

and for $r \leq M - r$

$$\liminf_{\alpha \rightarrow \infty} \frac{\mathbb{E}Z_{\hat{\pi}_r}}{\alpha} \gtrsim \sigma r \sqrt{M - r}. \quad (1.5)$$

(ii) Denote

$$Z_{\hat{\pi}_s}^s := \sup_{\tilde{\pi}_s \in \mathcal{S}_{M,r}} \|\tilde{\pi}_s (C_{\alpha,s} + E)\|_{S_2}^2 - \|\pi_s (C_{\alpha,s} + E)\|_{S_2}^2, \quad 1 \leq s < M,$$

where the singular value decomposition of $C_{\alpha,s}$ is given by $U \alpha \text{Id}_s V'$, $\alpha > 0$. Then it holds

$$\liminf_{\alpha \rightarrow \infty} \max_{s \in \{r, M-r\}} \mathbb{E}Z_{\hat{\pi}_s}^s \gtrsim \sigma^2 r (M - r). \quad (1.6)$$

(iii) Let $r=1$. There exists an $M_0 \in \mathbb{N}$, so that for all $\sigma^2 > 0$ and any $M \geq M_0$ it holds

$$\inf_{C \in \mathbb{R}^{M \times M}} \mathbb{E} Z_{\hat{\pi}_r} \gtrsim \mathbb{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2 \right). \quad (1.7)$$

(1.4), (1.6) and (1.7) indicate, that there does not exist a more favorable matrix than $C = 0$ in terms of accuracy of $\|\hat{\pi}_r X\|_{S_2}$ for $\|\pi_r X\|_{S_2}$. For $r = 1$ this statement is proven. (1.5) shows, that in general the upper bound $\sigma^2 r M (1 + \frac{\lambda_1}{\sigma \sqrt{M}})$ is unimprovable. Nevertheless it is possible to state a more refined upper bound, as seen in Theorem 1.

In this article we generalize Theorem 1 to all random matrices of centered i.i.d. entries with finite fourth moment. Our approach differs significantly from Rohde (2012). The key argument is an upper bound on $Z_{\hat{\pi}_r}$, whose randomness is totally determined by σ_1 . Concerning the results of Latała (2005), our generalization depends on both the variance and the fourth moments of the entries, since σ_1 does. In a broad sense we exploit the location $\hat{\pi}_r$ of the supremum of the process Z . The clue is, that given σ_1 , Z attains its supremum on a rather small S_2 -ball depending on σ_1 . Our upper bound on $Z_{\hat{\pi}_r}$ takes this into account.

The main result of this article is the following:

Theorem 3 (Universal upper bound)

Assume that the entries E_{ij} of the random matrix E have finite variance σ^2 and finite fourth moment m_4 . In this case the following inequality holds

$$\mathbb{E} Z_{\hat{\pi}_r} \lesssim r(M-r) \min(\text{I}, \text{II}, \text{III}), \quad (1.8)$$

where

$$\text{I} = \sigma^2 + \sqrt{m_4} + \frac{\lambda_1}{\sqrt{M}} (\sigma + \sqrt[4]{m_4}), \quad (1.9a)$$

$$\text{II} = \begin{cases} \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} (\sigma^2 + \sqrt{m_4}) & \text{if } \lambda_r > \lambda_{r+1}, \\ \infty & \text{if } \lambda_r = \lambda_{r+1}, \end{cases} \quad (1.9b)$$

$$\text{III} = \begin{cases} \frac{\lambda_1^2}{\lambda_r^2} (\sigma^2 + \sqrt{m_4}) + \sqrt{\frac{\lambda_1^2 \sum_{i=r+1}^{2r} \lambda_i^2}{r(M-r)\lambda_r^2}} (\sigma + \sqrt[4]{m_4}) & \text{if } \lambda_r > 0, \\ \infty & \text{if } \lambda_r = 0. \end{cases} \quad (1.9c)$$

This result is a generalization of Theorem 5.1 of Rohde (2012) (resp. Theorem 1 stated above). We give a brief discussion of this fact later.

The article is structured as follows. In the next section we introduce further notations. We give some elementary estimations on traces of certain matrices in the third section. Most of the results in this section are stated for deterministic matrices. In the fourth section a proof of Theorem 3 is given. Finally in the last section we give a further application of Proposition 1 of section 3. We derive intervals containing $\liminf_{M \rightarrow \infty} \lambda_1(C_M + E_M)$ and

$\limsup_{M \rightarrow \infty} \lambda_1(C_M + E_M)$ almost surely, where C_M is a deterministic $M \times M$ matrix and E_M is a $M \times M$ random matrix of i.i.d entries with variance $\sigma^2 M^{-1}$.

2 Preliminaries

We split Z into two subprocesses Z^1 and Z^2 given by

$$\begin{aligned} Z_{\tilde{\pi}_r}^1 &:= \|\tilde{\pi}_r C\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 + 2\text{tr}(E^T(\tilde{\pi}_r - \pi_r)C), \\ Z_{\tilde{\pi}_r}^2 &:= \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2. \end{aligned}$$

So it holds $Z = Z^1 + Z^2$. Further we denote by $\hat{\pi}_r^1$ a location of the supremum of Z^1 . $A \lesssim B$ means, that A is equal or less than B up to some universal constant. If $A \lesssim B$ and $B \lesssim A$, we write $A \sim B$. We denote the Schatten- p -norm, $1 \leq p \leq \infty$, on $\mathbb{R}^{M \times M}$ by $\|\cdot\|_{S_p}$. For $C \in \mathbb{R}^{M \times M}$ with singular values $\lambda_1 \geq \dots \geq \lambda_M$ the Schatten- p -norm of C is given by

$$\|C\|_{S_p} = \sqrt[p]{\sum_{i=1}^M \lambda_i^p} \text{ for } 1 \leq p \leq \infty \text{ and } \|C\|_{S_\infty} = \lambda_1.$$

In particular we will use the Hilbert-Schmidt norm $\|\cdot\|_{S_2}$ and the spectral norm $\|\cdot\|_{S_\infty}$. For the trace of a matrix $C \in \mathbb{R}^{M \times M}$ we write $\text{tr}(C)$ and for its transpose C^T . Moreover put $\Delta_r := \sum_{i=r+1}^{2r} \lambda_i^2$ and $r_M := r \wedge (M - r)$. The Euclidean sphere is denoted by S^{M-1} . For any set $B \subset \mathcal{S}_{M,r}$ we define $\overline{B} := \mathcal{S}_{M,r} \setminus B$. Lastly, $\lfloor x \rfloor$ is the smallest integer larger than $x \in \mathbb{R}$.

3 Estimation of traces involving differences of projection matrices

In this section we derive estimations of traces of certain matrices like those arising in the process Z . However, the results are stated in a quite general way and are phrased in a deterministic setting.

First recall some basic properties of orthogonal projections. By definition we have

$$\pi_r = \pi_r^T \text{ and } \pi_r = \pi_r \pi_r \text{ for } \pi_r \in \mathcal{S}_{M,r}.$$

Therefore every orthogonal projection π_r is positive-semidefinite. For $\pi_r^{(1)}, \pi_r^{(2)} \in \mathcal{S}_{M,r}$ with eigendecomposition $\pi_r^{(1)} = U \text{Id}_r U^T$ and $\pi_r^{(2)} = \tilde{U} \text{Id}_r \tilde{U}^T$ we get

$$\text{tr}(\pi_r^{(1)} \pi_r^{(2)}) = \text{tr}(U \text{Id}_r U^T \tilde{U} \text{Id}_r \tilde{U}^T) = \text{tr}(\tilde{U}^T U \text{Id}_r U^T \tilde{U} \text{Id}_r).$$

The matrix $P := \tilde{U}^T U \text{Id}_r U^T \tilde{U}$ is also an orthogonal projection. Since P is positive-semidefinite, the diagonal entries of P are nonnegative. It follows

$$\text{tr}(\pi_r^{(1)} \pi_r^{(2)}) = \text{tr}(P \text{Id}_r) = \sum_{i=1}^r P_{ii} \geq 0.$$

We conclude

$$\|\pi_r^{(2)} - \pi_r^{(1)}\|_{S_2} = \|(Id - \pi_r^{(1)}) - (Id - \pi_r^{(2)})\|_{S_2} \leq \sqrt{2r_M}.$$

Finally, note that by symmetry of $\pi_r^{(2)} - \pi_r^{(1)}$ we have

$$\begin{aligned} \|\pi_r^{(2)} - \pi_r^{(1)}\|_{S_\infty} &= \sup_{x \in S^{M-1}} |x^T (\pi_r^{(2)} - \pi_r^{(1)}) x| \\ &= \sup_{x \in S^{M-1}} \underbrace{|x^T \pi_r^{(2)} x|}_{\in [0,1]} - \underbrace{|x^T \pi_r^{(1)} x|}_{\in [0,1]} \leq 1. \end{aligned}$$

The next lemma provides a useful estimate to bound $\text{tr}(E^T(\tilde{\pi}_r - \pi_r)C)$ and $Z_{\tilde{\pi}_r}^2$.

Lemma 1

Let $\pi_r^{(1)}, \pi_r^{(2)} \in \mathcal{S}_{M,r}$ and $A, B \in \mathbb{R}^{M \times M}$, then the following inequality holds

$$\text{tr}(A^T(\pi_r^{(2)} - \pi_r^{(1)})B) \leq \sqrt{2r_M} \|A\|_{S_\infty} \|B\|_{S_\infty} \|\pi_r^{(2)} - \pi_r^{(1)}\|_{S_2}. \quad (3.1)$$

Proof. First, note that

$$\pi_r^{(2)} - \pi_r^{(1)} = \pi_r^{(2)} - \pi_r^{(2)}\pi_r^{(1)} + \pi_r^{(2)}\pi_r^{(1)} - \pi_r^{(1)} = \pi_r^{(2)}(Id - \pi_r^{(1)}) + (\pi_r^{(2)} - Id)\pi_r^{(1)}.$$

From the proof of Proposition 8.1 in Rohde (2012) we get

$$\|(Id - \pi_r^{(2)})\pi_r^{(1)}\|_{S_2} = \|\pi_r^{(2)}(Id - \pi_r^{(1)})\|_{S_2} = \frac{1}{\sqrt{2}} \|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_2}. \quad (3.2)$$

By Cauchy-Schwarz inequality follows

$$\begin{aligned} &\text{tr}(A^T(\pi_r^{(2)} - \pi_r^{(1)})B) \\ &= \text{tr}(A^T \pi_r^{(2)}(Id - \pi_r^{(1)})B) + \text{tr}(A^T(\pi_r^{(2)} - Id)\pi_r^{(1)}B) \\ &\leq \left(\|BA^T \pi_r^{(2)}\|_{S_2} \wedge \|(Id - \pi_r^{(1)})BA^T\|_{S_2} \right) \|\pi_r^{(2)}(Id - \pi_r^{(1)})\|_{S_2} \\ &\quad + \left(\|\pi_r^{(1)}BA^T\|_{S_2} \wedge \|BA^T(Id - \pi_r^{(2)})\|_{S_2} \right) \|(\pi_r^{(2)} - Id)\pi_r^{(1)}\|_{S_2} \\ &\leq \frac{1}{\sqrt{2}} \sqrt{r_M} \sqrt{\|BA^T AB^T\|_{S_\infty}} \|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_2} \\ &\quad + \frac{1}{\sqrt{2}} \sqrt{r_M} \sqrt{\|BA^T AB^T\|_{S_\infty}} \|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_2} \\ &\leq \sqrt{2r_M} \|A\|_{S_\infty} \|B\|_{S_\infty} \|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_2}. \end{aligned}$$

□

The statement of the lemma is optimal to the effect, that in case of $M \geq 2r$, orthonormal vectors $u_1, \dots, u_r, \tilde{u}_1, \dots, \tilde{u}_r$ and matrices

$$\begin{aligned}\pi_r^{(1)} &= \sum_{i=1}^r u_i u_i^T, \quad \pi_r^{(2)} = \sum_{i=1}^r (\sqrt{1-\alpha^2} u_i + \alpha \tilde{u}_i)(\sqrt{1-\alpha^2} u_i + \alpha \tilde{u}_i)^T, \\ A &= \mu \text{Id}, \quad B = \nu \left(\pi_r^{(1)} - \pi_r^{(2)} \right),\end{aligned}$$

the left-hand side of the inequality attains the upper bound for any real numbers $0 \leq \alpha \leq 1$, $\mu, \nu > 0$. We give a brief computation:

$$\begin{aligned}\text{tr}(A^T(\pi_r^{(1)} - \pi_r^{(2)})B) &= \mu\nu \text{tr} \left(\left(\pi_r^{(1)} - \sum_{i=1}^r (\sqrt{1-\alpha^2} u_i + \alpha \tilde{u}_i)(\sqrt{1-\alpha^2} u_i + \alpha \tilde{u}_i)^T \right) \right. \\ &\quad \times \left. \left(\pi_r^{(1)} - \sum_{i=1}^r (\sqrt{1-\alpha^2} u_i + \alpha \tilde{u}_i)(\sqrt{1-\alpha^2} u_i + \alpha \tilde{u}_i)^T \right) \right) \\ &= \mu\nu \left(2r - 2\text{tr} \left(\pi_r^{(1)} \pi_r^{(2)} \right) \right) \\ &= \mu\nu \left(2r - 2r(1-\alpha^2) \right) \\ &= \sqrt{2r} \mu \nu \alpha^2 \sqrt{2r}.\end{aligned}$$

So it remains to show, that $\|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_2} = \alpha\sqrt{2r}$ and $\|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_\infty} = \alpha$. The first equation is obvious concerning the previous calculation, since

$$\|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_2} = \sqrt{\text{tr} \left(\left(\pi_r^{(1)} - \pi_r^{(2)} \right) \left(\pi_r^{(1)} - \pi_r^{(2)} \right) \right)} = \alpha\sqrt{2r}.$$

To prove the second equation, one can check, that α and $-\alpha$ are the only non-zero eigenvalues of $\pi_r^{(1)} - \pi_r^{(2)}$ and their eigenspaces are given by

$$W_\alpha = \text{span} \left\{ \sqrt{\frac{1+\alpha}{2}} u_i - \sqrt{\frac{1-\alpha}{2}} \tilde{u}_i \mid i = 1, \dots, r \right\}$$

and

$$W_{-\alpha} = \text{span} \left\{ \sqrt{\frac{1-\alpha}{2}} u_i + \sqrt{\frac{1+\alpha}{2}} \tilde{u}_i \mid i = 1, \dots, r \right\}.$$

Since $\pi_r^{(1)} - \pi_r^{(2)}$ is symmetric, it holds

$$\|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_\infty} = \max(\lambda_{\max}(\pi_r^{(1)} - \pi_r^{(2)}), |\lambda_{\min}(\pi_r^{(1)} - \pi_r^{(2)})|) = \alpha.$$

As Z has a negative drift, Lemma 1 is not useful to bound $\|\tilde{\pi}_r C\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2$. Therefore the next lemma gives an estimate on the drift term. It is significant for our subsequent computations, that the distance $\|\tilde{\pi}_r - \pi_r\|_{S_2}$ influences the drift term rather squared than linearly.

Lemma 2

(i) For any $\tilde{\pi}_r \in \mathcal{S}_{M,r}$ the following inequality holds

$$\|\tilde{\pi}_r C\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 \leq -\frac{1}{2} (\lambda_r^2 - \lambda_{r+1}^2) \|\tilde{\pi}_r - \pi_r\|_{S_2}^2. \quad (3.3)$$

(ii) For any $\tilde{\pi}_r \in \mathcal{S}_{M,r}$, such that $\|\tilde{\pi}_r - \pi_r\|_{S_2} \geq \lambda_r^{-1} \sqrt{2\Delta_r}$, we have

$$\|\tilde{\pi}_r C\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 \leq -\frac{1}{2} \lambda_r^2 \|\tilde{\pi}_r - \pi_r\|_{S_2}^2 + \Delta_r. \quad (3.4)$$

Proof. Both inequalities are clear for $\lambda_r = 0$. Hence assume $\lambda_r > 0$.

(i): It is obvious in case of $\lambda_r = \lambda_{r+1}$ or $\tilde{\pi}_r = \pi_r$, so assume moreover $\lambda_r > \lambda_{r+1}$ and $\tilde{\pi}_r \neq \pi_r$. Now (i) follows easily from Proposition 8.1 in Rohde (2012) combined with an epsilon of room argument. Let $\varepsilon > 0$ and set

$$\|\tilde{\pi}_r - \pi_r\|_{S_2} > \|\tilde{\pi}_r - \pi_r\|_{S_2} - \varepsilon = \sqrt{\frac{1}{\lambda_r^2 - \lambda_{r+1}^2}} \sqrt{2\delta}.$$

Then from statement (8.1) in the mentioned proposition we get

$$\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 > \delta = \frac{1}{2} (\lambda_r^2 - \lambda_{r+1}^2) (\|\pi_r - \tilde{\pi}_r\|_{S_2} - \varepsilon)^2.$$

By multiplication by -1 and taking the limit $\varepsilon \rightarrow 0$ (i) is proved.

(ii): This part follows analogous to (i). \square

Now we derive an upper bound on $Z_{\hat{\pi}_r}^1$, which will be useful to estimate the expectation of $Z_{\hat{\pi}_r}$. In a certain way the upper bound regards the location of $\hat{\pi}_r^1$.

Proposition 1

For the supremum $Z_{\hat{\pi}_r}^1$ of the process Z^1 we have $Z_{\hat{\pi}_r}^1 \leq Y$ with

$$Y := \min (\text{I}', \text{II}', \text{III}') , \quad (3.5)$$

where

$$\text{I}' := 4r_M \lambda_1 \sigma_1, \quad (3.6a)$$

$$\text{II}' := \begin{cases} 4r_M \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \sigma_1^2 & \text{if } \lambda_r > \lambda_{r+1}, \\ \infty & \text{if } \lambda_r = \lambda_{r+1}, \end{cases} \quad (3.6b)$$

$$\text{III}' := \begin{cases} \max \left(4\sqrt{r_M \Delta_r} \frac{\lambda_1}{\lambda_r} \sigma_1, 8r_M \frac{\lambda_1^2}{\lambda_r^2} \sigma_1^2 \right) & \text{if } \lambda_r > 0, \\ \infty & \text{if } \lambda_r = 0. \end{cases} \quad (3.6c)$$

Proof. We prove for I', II' and III' , that $Z_{\tilde{\pi}_r}^1$ is less or equal to each one.

$Z_{\tilde{\pi}_r}^1 \leq I'$: Since $\|\tilde{\pi}_r C\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 \leq 0$, we get by Lemma 1 for any $\tilde{\pi}_r \in \mathcal{S}_{M,r}$

$$Z_{\tilde{\pi}_r}^1 \leq 2\sqrt{2r_M}\sigma_1\lambda_1\|\tilde{\pi}_r - \pi_r\|_{S_2} \leq 4r_M\sigma_1\lambda_1. \quad (3.7)$$

As $4r_M\sigma_1\lambda_1$ does not depend on $\tilde{\pi}_r$, $Z_{\tilde{\pi}_r}^1 \leq I'$ holds.

$Z_{\tilde{\pi}_r}^1 \leq II'$: Assume $\lambda_r > \lambda_{r+1}$. We obtain by Lemma 1 and 2(i) for any $\tilde{\pi}_r \in \mathcal{S}_{M,r}$

$$\begin{aligned} Z_{\tilde{\pi}_r}^1 &= \|\tilde{\pi}_r C\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 + 2\text{tr}(E^T(\tilde{\pi}_r - \pi_r)C) \\ &\leq -\frac{1}{2}(\lambda_r^2 - \lambda_{r+1}^2)\|\tilde{\pi}_r - \pi_r\|_{S_2}^2 + 2\sqrt{2r_M}\sigma_1\lambda_1\|\tilde{\pi}_r - \pi_r\|_{S_2} \end{aligned}$$

Then maximizing the right-hand side of the equation

$$Z_{\tilde{\pi}_r}^1 \leq -\frac{1}{2}(\lambda_r^2 - \lambda_{r+1}^2)\|\tilde{\pi}_r - \pi_r\|_{S_2}^2 + 2\sqrt{2r_M}\sigma_1\lambda_1\|\tilde{\pi}_r - \pi_r\|_{S_2}$$

over all $\tilde{\pi}_r \in \mathcal{S}_{M,r}$ provides the claim.

$Z_{\tilde{\pi}_r}^1 \leq III'$: Assume $\lambda_r > 0$. In order to prove the last bound we split $\mathcal{S}_{M,r}$ into two sets and take the supremum of Z^1 on this sets separately. We define

$$B_{III'} := \{\tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\tilde{\pi}_r - \pi_r\|_{S_2} < \lambda_r^{-1}\sqrt{2\Delta_r}\}. \quad (3.8)$$

It holds

$$Z_{\tilde{\pi}_r}^1 = \max \left(\sup_{\tilde{\pi}_r \in B_{III'}} Z_{\tilde{\pi}_r}^1, \sup_{\tilde{\pi}_r \in \overline{B}_{III'}} Z_{\tilde{\pi}_r}^1 \right). \quad (3.9)$$

For the first expression in the maximum of (3.9) we get analogous to the proof of $Z_{\tilde{\pi}_r}^1 \leq I'$:

$$\sup_{\tilde{\pi}_r \in B_{III'}} Z_{\tilde{\pi}_r}^1 \leq \sup_{\tilde{\pi}_r \in B_{III'}} 2\sqrt{2r_M}\sigma_1\lambda_1\|\tilde{\pi}_r - \pi_r\|_{S_2} \leq 4\sqrt{r_M\Delta_r}\frac{\lambda_1}{\lambda_r}\sigma_1. \quad (3.10)$$

It remains to bound the second expression in the maximum of (3.9). By Lemma 1 again and by Lemma 2(ii) follows for any $\tilde{\pi}_r \in \overline{B}_{III'}$

$$Z_{\tilde{\pi}_r}^1 \leq -\frac{1}{2}\lambda_r^2\|\tilde{\pi}_r - \pi_r\|_{S_2}^2 + \Delta_r + 2\sqrt{2r_M}\sigma_1\lambda_1\|\tilde{\pi}_r - \pi_r\|_{S_2}. \quad (3.11)$$

The right-hand side attains its global maximum on

$$\{\tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\tilde{\pi}_r - \pi_r\|_{S_2} = 2\sqrt{2r_M}\sigma_1\frac{\lambda_1}{\lambda_r^2} \wedge \sqrt{2r_M}\}. \quad (3.12)$$

If $2\sqrt{2r_M}\sigma_1\frac{\lambda_1}{\lambda_r} < \sqrt{2\Delta_r}$, then it holds

$$\{\tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\tilde{\pi}_r - \pi_r\|_{S_2} = 2\sqrt{2r_M}\sigma_1\frac{\lambda_1}{\lambda_r^2} \wedge \sqrt{2r_M}\} \cap \overline{B}_{III'} = \emptyset.$$

In this case due to reasons of monotonicity the right-hand side of (3.11) restricted to $\{\tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\tilde{\pi}_r - \pi_r\|_{S_2} \geq \lambda_r^{-1} \sqrt{2\Delta_r}\}$ attains its maximum on $\{\tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\tilde{\pi}_r - \pi_r\|_{S_2} = \lambda_r^{-1} \sqrt{2\Delta_r}\}$. So we have

$$\mathbb{1}_{\{\sqrt{2\Delta_r} > 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} Z_{\tilde{\pi}_r}^1 \leq \mathbb{1}_{\{\sqrt{2\Delta_r} > 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} 4\sqrt{r_M \Delta_r} \frac{\lambda_1}{\lambda_r} \sigma_1. \quad (3.13)$$

Otherwise by (3.12) follows

$$\begin{aligned} \mathbb{1}_{\{\sqrt{2\Delta_r} \leq 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} Z_{\tilde{\pi}_r}^1 &\leq \mathbb{1}_{\{\sqrt{2\Delta_r} \leq 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} \left(4r_M \frac{\lambda_1^2}{\lambda_r^2} \sigma_1^2 + \Delta_r \right) \\ &\leq \mathbb{1}_{\{\sqrt{2\Delta_r} \leq 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} 8r_M \frac{\lambda_1^2}{\lambda_r^2} \sigma_1^2. \end{aligned} \quad (3.14)$$

By (3.13) and (3.14) we get

$$\begin{aligned} \sup_{\tilde{\pi}_r \in \tilde{B}_{III'}} Z_{\tilde{\pi}_r}^1 &= \sup_{\tilde{\pi}_r \in \tilde{B}_{III'}} \mathbb{1}_{\{\sqrt{2\Delta_r} > 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} Z_{\tilde{\pi}_r}^1 + \mathbb{1}_{\{\sqrt{2\Delta_r} \leq 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} Z_{\tilde{\pi}_r}^1 \\ &\leq \mathbb{1}_{\{\sqrt{2\Delta_r} > 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} 4\sqrt{r_M \Delta_r} \frac{\lambda_1}{\lambda_r} \sigma_1 + \mathbb{1}_{\{\sqrt{2\Delta_r} \leq 2\sqrt{2r_M} \frac{\lambda_1}{\lambda_r} \sigma_1\}} 8r_M \frac{\lambda_1^2}{\lambda_r^2} \sigma_1^2 \\ &\leq \max \left(4\sqrt{r_M \Delta_r} \frac{\lambda_1}{\lambda_r} \sigma_1, 8r_M \frac{\lambda_1^2}{\lambda_r^2} \sigma_1^2 \right). \end{aligned} \quad (3.15)$$

Finally combining (3.10) and (3.15) yields $Z_{\tilde{\pi}_r}^1 \leq III'$. \square

4 Proof of Theorem 3

Now we are ready to prove Theorem 3. We first state a result by Latała (2005) in simplified terms.

Theorem 4 (Latała (2005))

For any random matrix E of centered i.i.d. entries with variance σ^2 and fourth moment m_4 the following inequality holds

$$\mathbb{E} \sigma_1^2 \lesssim M \left(\sigma^2 + \sqrt{m_4} \right). \quad (4.1)$$

Note that the original result is phrased for the expectation of σ_1 and not of σ_1^2 , but actually the proof includes statement (4.1) under the additional assumption, that the distribution of the entries is symmetric. So one only has to adjust the symmetrization argument from $\mathbb{E} \|E\|_{S_\infty}$ to $\mathbb{E} \|E\|_{S_\infty}^2$. Therefore let \tilde{E} be an independent copy of E and ε a $M \times M$ -random matrix of i.i.d. Rademacher variables, which is independent of E and \tilde{E} . Then we

obtain by Jensen's inequality for conditional expectation

$$\begin{aligned}
\mathbb{E}\sigma_1^2 &= \mathbb{E}\|E\|_{S_\infty}^2 \\
&= \mathbb{E}\left(\sup_{x,y \in S^{M-1}} x^T E y\right)^2 \\
&= \mathbb{E}\sup_{x,y \in S^{M-1}} (x^T E y)^2 \\
&= \mathbb{E}\sup_{x,y \in S^{M-1}} (x^T E y - \mathbb{E}x^T \tilde{E}y)^2 \\
&= \mathbb{E}\sup_{x,y \in S^{M-1}} \left(\mathbb{E}(x^T E y - x^T \tilde{E}y | E)\right)^2 \\
&\leq \mathbb{E}\sup_{x,y \in S^{M-1}} \mathbb{E}\left((x^T E y - x^T \tilde{E}y)^2 | E\right) \\
&\leq \mathbb{E}\mathbb{E}\left(\sup_{x,y \in S^{M-1}} (x^T E y - x^T \tilde{E}y)^2 | E\right) \\
&= \mathbb{E}\sup_{x,y \in S^{M-1}} (x^T E y - x^T \tilde{E}y)^2 \\
&= \mathbb{E}\|E - \tilde{E}\|_{S_\infty}^2 \\
&= \mathbb{E}\|\varepsilon \cdot (E - \tilde{E})\|_{S_\infty}^2 \\
&\leq \mathbb{E}\left(\|\varepsilon \cdot E\|_{S_\infty} + \|\varepsilon \cdot \tilde{E}\|_{S_\infty}\right)^2 \\
&\leq \mathbb{E}\left(2\|\varepsilon \cdot E\|_{S_\infty}^2 + 2\|\varepsilon \cdot \tilde{E}\|_{S_\infty}^2\right) \\
&\leq 4\mathbb{E}\|\varepsilon \cdot E\|_{S_\infty}^2,
\end{aligned}$$

where \cdot denotes the Hadamard product. Since E_{ij} and $\varepsilon_{ij}E_{ij}$ have the same variance and fourth moment, statement (4.1) holds without the symmetry assumption on the distribution of the entries.

Proof of Theorem 3. Beforehand note, that by distinguishing the cases $r < \frac{M}{2}$ and $r \geq \frac{M}{2}$ it holds

$$r(M - r) \leq r_M M \leq 2r(M - r).$$

Now we commence the proof of Theorem 3. First we get

$$\mathbb{E}Z_{\tilde{\pi}_r} \leq \mathbb{E}Z_{\tilde{\pi}_r}^1 + \mathbb{E}\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} Z_{\tilde{\pi}_r}^2. \quad (4.2)$$

Applying Theorem 4 to the second summand yields

$$\begin{aligned}
\mathbb{E}\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} Z_{\tilde{\pi}_r}^2 &\leq 2r_M \mathbb{E}\sigma_1^2 \lesssim r_M M \left(\sigma^2 + \sqrt{m^4}\right) \\
&\lesssim r(M - r) \left(\sigma^2 + \sqrt{m^4}\right). \quad (4.3)
\end{aligned}$$

Therefore

$$\mathbb{E} \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} Z_{\tilde{\pi}_r}^2 \lesssim r(M-r) \min(\text{I}, \text{II}, \text{III}). \quad (4.4)$$

So it remains to prove, that

$$\mathbb{E} Z_{\tilde{\pi}_r}^1 \lesssim r(M-r) \min(\text{I}, \text{II}, \text{III}). \quad (4.5)$$

By Proposition 1, monotonicity of integral and Theorem 4 we get

$$\begin{aligned} \mathbb{E} Z_{\tilde{\pi}_r}^1 &\leq \mathbb{E} Y \\ &\leq \min(\text{EI}', \text{EII}', \text{EIII}') \\ &\leq \min \left(4r_M \lambda_1 \mathbb{E} \sigma_1, 4r_M \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \mathbb{E} \sigma_1^2, \mathbb{E} \max \left(4\sqrt{r_M} \Delta_r \frac{\lambda_1}{\lambda_r} \sigma_1, 8r_M \frac{\lambda_1^2}{\lambda_r^2} \sigma_1^2 \right) \right) \\ &\leq \min \left(4r_M \lambda_1 \mathbb{E} \sigma_1, 4r_M \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \mathbb{E} \sigma_1^2, 4\sqrt{r_M} \Delta_r \frac{\lambda_1}{\lambda_r} \mathbb{E} \sigma_1 + 8r_M \frac{\lambda_1^2}{\lambda_r^2} \mathbb{E} \sigma_1^2 \right) \\ &\lesssim r(M-r) \min \left(\frac{\lambda_1}{\sqrt{M}} (\sigma + \sqrt[4]{m_4}), \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} (\sigma^2 + \sqrt[4]{m_4}), \right. \\ &\quad \left. \frac{\lambda_1^2}{\lambda_r^2} (\sigma^2 + \sqrt{m_4}) + \sqrt{\frac{\lambda_1^2 \sum_{i=r+1}^{2r} \lambda_i^2}{r(M-r)\lambda_r^2}} (\sigma + \sqrt[4]{m_4}) \right) \\ &\lesssim r(M-r) \min(\text{I}, \text{II}, \text{III}). \end{aligned}$$

□

As mentioned in the introduction, this result is a generalization of Theorem 5.1 of Rohde (2012). To check this consider the case, where E is a Gaussian matrix and $r \leq M-r$. Since the fourth moment of a centered Gaussian random variable is given by $3\sigma^4$, the right-hand side of inequality (1.8) may be rewritten as

$$\sigma^2 r M \min \left(1 + \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}, \frac{\lambda_1^2}{\lambda_r^2} + \sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}} \right),$$

where the constant in (1.8) is now specific to Gaussian matrices. So we have to show, that

$$\begin{aligned} &\sigma^2 r M \left(\min \left(\frac{\lambda_1^2}{\lambda_r^2}, 1 + \frac{\lambda_1}{\sigma \sqrt{M}} \right) + \min \left(\sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \right) \right) \\ &\lesssim \sigma^2 r M \min \left(1 + \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}, \frac{\lambda_1^2}{\lambda_r^2} + \sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}} \right) \\ &\lesssim \sigma^2 r M \left(\min \left(\frac{\lambda_1^2}{\lambda_r^2}, 1 + \frac{\lambda_1}{\sigma \sqrt{M}} \right) + \min \left(\sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \right) \right). \end{aligned}$$

This follows by (4.6) and (4.7) in the next computation

$$\begin{aligned}
& \sigma^2 r M \left(\min \left(\frac{\lambda_1^2}{\lambda_r^2}, 1 + \frac{\lambda_1}{\sigma \sqrt{M}} \right) + \min \left(\sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \right) \right) \\
& \leq \sigma^2 r M \min \left(1 + \frac{\lambda_1}{\sigma \sqrt{M}} + \sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}}, \right. \\
& \quad \left. \frac{\lambda_1^2}{\lambda_r^2} + \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}, \frac{\lambda_1^2}{\lambda_r^2} + \sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}} \right) \\
& \leq 2\sigma^2 r M \min \left(1 + \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}, \frac{\lambda_1^2}{\lambda_r^2} + \sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}} \right) \quad (4.6) \\
& \leq 2\sigma^2 r M \min \left(1 + \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2} + \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}, \frac{\lambda_1^2}{\lambda_r^2} + \sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}} \right) \\
& = 2\sigma^2 r M \min \left(1 + \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2} + \min \left(\frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}, \sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}} \right) \right) \\
& \leq 2\sigma^2 r M \left(\min \left(\frac{\lambda_1^2}{\lambda_r^2}, 1 + \frac{\lambda_1}{\sigma \sqrt{M}} \right) \right. \\
& \quad \left. + \min \left(\sqrt{\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2}} \frac{\lambda_1}{\sigma \sqrt{M}}, \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \right) \right). \quad (4.7)
\end{aligned}$$

The last line arises by the simple observation, that $\min(a, b + c) \leq \min(a, b) + c$ for any $a, b \in \mathbb{R}$ and $c \geq 0$.

5 Application: Localizing the largest singular value of a deformed random matrix

As a further application of Proposition 1 we take a classical view on random matrices. Hence, let $(E_{ij})_{i,j \in \mathbb{N}}$ be a doubly indexed sequence of centered i.i.d. random variables with variance σ^2 and finite fourth moment. By (E_M) we denote the sequence of $M \times M$ random matrices $1/\sqrt{M}(E_{ij})_{i,j \leq M}$. (C_M) is a sequence of deterministic $M \times M$ matrices. Assume furthermore that the first and the second singular values $\lambda_1(C_M)$ and $\lambda_2(C_M)$ converge to some real numbers $\lambda_1 > \lambda_2 \geq 0$ for $M \rightarrow \infty$. We specify an interval containing $\liminf_{M \rightarrow \infty} \lambda_1(C_M + E_M)$ and $\limsup_{M \rightarrow \infty} \lambda_1(C_M + E_M)$ almost surely.

Corollary 1

Under the former notations and assumptions let $(u_{i1})_{i \in \mathbb{N}}$ be a sequence of real numbers,

such that

$$\left(\sum_{i=1}^M u_{i1}^2 \right)^{-\frac{1}{2}} (u_{11}, \dots, u_{M1})^T$$

is the left singular vector of C_M corresponding to the largest singular value. If there exist $\beta > 1, \beta' > 0$ with

$$\frac{\sum_{i=B}^M u_{i1}^2}{\sum_{i=1}^M u_{i1}^2} \lesssim \frac{1}{M^{\beta'}} \text{ for all } M \in \mathbb{N}, \text{ where } B = \lfloor (M^{\frac{1}{\beta}} - 1)^\beta \rfloor, \quad (5.1)$$

then it holds a.s.

$$\begin{aligned} \sqrt{\lambda_1^2 + \sigma^2} &\leq \liminf_{M \rightarrow \infty} \lambda_1(C^M + E^M) \\ &\leq \limsup_{M \rightarrow \infty} \lambda_1(C^M + E^M) \leq \sqrt{\lambda_1^2 + 4\sigma^2 + 16\sigma^2 \frac{\lambda_1^2}{\lambda_1^2 - \lambda_2^2}}. \end{aligned} \quad (5.2)$$

Thus, if in the large amplitude regime the values λ_1 and λ_2 are well-separated, then the largest singular value of $C_M + E_M$ is typically close to λ_1 but larger. This result can be seen complementary to Benaych-Georges and Nadakuditi (2012). They consider finite rank perturbations of a sequence of random matrices (X_M) , where X_M is a $M \times N$ -matrix. Under certain assumptions they show an almost sure convergence of the largest singular values in the limit $M, N_M \rightarrow \infty$. Since we only make assumptions on the first two singular values and the first left singular vectors of the perturbation matrices (C_M) , the limit $\lim_{M \rightarrow \infty} \lambda_1(C_M + E_M)$ does not exist in general. However, under the condition $\lambda_1 - \lambda_2 \gg 10\sigma^2$, there exists a small interval containing $\liminf_{M \rightarrow \infty} \lambda_1(C_M + E_M)$ and $\limsup_{M \rightarrow \infty} \lambda_1(C_M + E_M)$ almost surely. Note that if $\lambda_1 - \lambda_2 \geq 4\sigma$, then the upper bound in Corollary 1 is already better than the obvious bound $\lambda_1 + 2\sigma$ on $\limsup_{M \rightarrow \infty} \lambda_1(C_M + E_M)$.

Before we prove Corollary 1, let us give two examples of sequences $(u_i)_{i \in \mathbb{N}}$ satisfying condition (5.1):

- All but finitely many u_i 's are zero.
- The sequence is bounded and bounded away from zero.

Proof of Corollary 1. Now we give a computation of (5.2). For this purpose we make a slight abuse of notations. We write C and E for the matrices C_M and E_M . Further let (v_{11}, \dots, v_{M1}) be the right singular vector of C_M corresponding to $\lambda_1(C_M)$. Consider the lower bound on $\liminf_{M \rightarrow \infty} \lambda_1(C_M + E_M)$:

$$\begin{aligned} \lambda_1(C + E) &\geq \|\pi_1(C + E)\|_{S_2} = \sqrt{\lambda_1(C)^2 + 2\text{tr}(C^T \pi_1 E) + \|\pi_1 E\|_{S_2}^2} \\ &= \sqrt{\lambda_1(C)^2 + 2\lambda_1(C) \sum_{i,j=1}^M \frac{v_{i1} u_{j1}}{(\sum_{i=1}^M u_{i1}^2)^{1/2}} E_{ji} + \sum_{i,j,k=1}^M \frac{u_{i1} u_{j1}}{\sum_{i=1}^M u_{i1}^2} E_{ik} E_{jk}}. \end{aligned}$$

By a SLLN of Thrum (1987) we get

$$\sum_{i,j=1}^M \frac{v_{i1}u_{j1}}{(\sum_{i=1}^M u_{i1}^2)^{1/2}} E_{ji} \xrightarrow{a.s.} 0.$$

Moreover, by the subsequent theorem follows

$$\sum_{i,j,k=1}^M \frac{u_{i1}u_{j1}}{\sum_{i=1}^M u_{i1}^2} E_{ik} E_{jk} \xrightarrow{a.s.} \sigma^2.$$

So it holds

$$\sqrt{\lambda_1^2 + \sigma^2} \leq \liminf_{M \rightarrow \infty} \lambda_1(C_M + E_M) \text{ a.s.}$$

It remains to prove the upper bound. It holds

$$\begin{aligned} \lambda_1(C + E) &= \sqrt{\|\hat{\pi}_1(C + E)\|_{S_2}^2 - \|\pi_1(C + E)\|_{S_2}^2 + \|\pi_1(C + E)\|_{S_2}^2} \\ &\leq \sqrt{\Pi' + \sigma_1^2 + \lambda_1(C)^2 + 2\lambda_1(C) \sum_{i,j=1}^M \frac{v_{i1}u_{j1}}{(\sum_{i=1}^M u_{i1}^2)^{1/2}} E_{ji}} \\ &= \sqrt{4 \frac{\lambda_1(C)^2}{\lambda_1(C)^2 - \lambda_2(C)^2} \sigma_1^2 + \sigma_1^2 + \lambda_1(C)^2 + 2\lambda_1(C) \sum_{i,j=1}^M \frac{v_{i1}u_{j1}}{(\sum_{i=1}^M u_{i1}^2)^{1/2}} E_{ji}}. \end{aligned}$$

By results of Bai, Krishnaiah, and Yin (1988) and Bai, Silverstein, and Yin (1988) we know that the fourth-moment condition on the entries of E is necessary and sufficient for the almost sure convergence of σ_1 to 2σ . Applying this to the last line of the computation yields the desired claim. \square

We close this article by a strong law of large numbers for empirical covariance matrices. In this result the empirical covariance matrix is considered as a quadratic form.

Theorem 5 (SLLN for empirical covariance matrices)

Let $(u_i)_{i \in \mathbb{N}}$ be a sequence of real numbers, such that there exist $\beta > 1, \beta' > 0$ with

$$\frac{\sum_{i=B}^M u_i^2}{\sum_{i=1}^M u_i^2} \lesssim \frac{1}{M^{\beta'}} \text{ for all } M \in \mathbb{N}, \text{ where } B = \lfloor (M^{\frac{1}{\beta}} - 1)^\beta \rfloor. \quad (5.3)$$

Furthermore let $(E_{ij})_{i,j \in \mathbb{N}}$ be a doubly indexed sequence of centered i.i.d. random variables with variance σ^2 and finite fourth moment. By (E_M) we denote the sequence of $M \times M$ random matrices $1/\sqrt{M}(E_{ij})_{i,j \leq M}$. Then we have

$$\tilde{u}_M^T E_M E_M^T \tilde{u}_M \xrightarrow{a.s.} \sigma^2, \quad (5.4)$$

where

$$\tilde{u}_M := \left(\sum_{i=1}^M u_i^2 \right)^{-\frac{1}{2}} (u_1, \dots, u_M)^T.$$

Proof. Set

$$Z_M := \tilde{u}_M^T E_M E_M^T \tilde{u}_M = \frac{1}{M \sum_{i=1}^M u_i^2} \sum_{k=1}^M \sum_{j=1}^M u_i u_j E_{ik} E_{jk}.$$

Therefore Z_M is the sum of W_M and $2X_M$ given by

$$\begin{aligned} W_M &:= \frac{1}{M \sum_{i=1}^M u_i^2} \sum_{k=1}^M \sum_{i=1}^M u_i^2 E_{ik}^2, \\ X_M &:= \frac{1}{M \sum_{i=1}^M u_i^2} \sum_{k=1}^M \sum_{i=1}^M \sum_{j=1}^{i-1} u_i u_j E_{ik} E_{jk}. \end{aligned}$$

First we show, that W_M converges to σ^2 almost surely. This part is an adaption of some arguments of the classical strong law of large numbers (cf. Etemadi (1981)). Here we do not even need truncation arguments, since the entries E_{ij} have finite fourth moments. By the Borel-Cantelli Lemma we get for $k_n := \lfloor n^\beta \rfloor$, $n \in \mathbb{N}$, that (W_{k_n}) converges to σ^2 almost surely. By monotonicity of $(M \sum_{i=1}^M u_i^2 W_M)$, condition (5.3) and picking $k_n \leq M \leq k_{n+1}$, it follows

$$\sigma^2 \leq \liminf_{M \rightarrow \infty} W_M \leq \limsup_{M \rightarrow \infty} W_M \leq \sigma^2 \text{ a.s.}$$

Now consider (X_M) . Let k_n be as before. Then again by the Borel-Cantelli Lemma (X_{k_n}) converges almost surely to 0. For any $M \in \mathbb{N}$ pick $n \in \mathbb{N}$, so that $k_n \leq M \leq k_{n+1}$. We have

$$\begin{aligned} |X_M| &\leq \frac{1}{M \sum_{i=1}^M u_i^2} \left| \sum_{k=1}^{k_n} \sum_{i=1}^{k_n} \sum_{j=1}^{i-1} u_i u_j E_{ik} E_{jk} \right| \\ &\quad + \frac{1}{M \sum_{i=1}^M u_i^2} \left| \sum_{k=1}^{k_n} \sum_{i=k_n+1}^M \sum_{j=1}^{i-1} u_i u_j E_{ik} E_{jk} \right| \\ &\quad + \frac{1}{M \sum_{i=1}^M u_i^2} \left| \sum_{k=k_n+1}^M \sum_{i=1}^M \sum_{j=1}^{i-1} u_i u_j E_{ik} E_{jk} \right| \end{aligned}$$

Clearly, the first and the last term go to zero almost surely. It remains to prove, that

$$V_M := \frac{1}{M \sum_{i=1}^M u_i^2} \sum_{k=1}^{k_n} \sum_{i=k_n+1}^M \sum_{j=1}^{i-1} u_i u_j E_{ik} E_{jk} \rightarrow 0 \text{ a.s.}$$

Therefore we estimate $\text{Var}(V_M)$:

$$\begin{aligned} \text{Var}(V_M) &\leq \frac{\sigma^4}{M \left(\sum_{i=1}^M u_i^2 \right)^2} \sum_{i=k_n+1}^M u_{1i}^2 \cdot \sum_{i=1}^{k_n} u_{1i}^2 \\ &\leq \frac{\sigma^4}{M} \cdot \frac{\sum_{i=k_n+1}^M u_{1i}^2}{\sum_{i=1}^M u_i^2} \\ &\leq \frac{\sigma^4}{M} \cdot \frac{\sum_{i=B}^M u_{1i}^2}{\sum_{i=1}^M u_i^2} \\ &\lesssim \frac{\sigma^4}{M^{1+\beta'}}. \end{aligned}$$

By the Borel-Cantelli Lemma and Chebyshev's inequality again we conclude the desired claim. \square

If only finitely many entries u_i are non-zero, then the almost sure convergence follows directly from the classical strong law of large numbers, since

$$Z_M = \frac{1}{M} \sum_{k=1}^M \left(\sum_{i=1}^M \frac{u_i}{\sqrt{\sum_{i=1}^M u_i^2}} E_{ik} \right)^2$$

and the summands

$$\left(\sum_{i=1}^M \frac{u_i}{\sqrt{\sum_{i=1}^M u_i^2}} E_{ik} \right)^2, \quad M \geq M_0,$$

are i.i.d. for M_0 large enough.

Acknowledgments

I am grateful to my Ph.D. advisor, Angelika Rohde, for her encouragement and bringing this topic to my attention.

References

- Bai, Z. D., Krishnaiah, P. R., and Yin, Y. Q. (1988). “On the limit of the largest eigenvalue of the large dimensional sample covariance matrix”. In: *Probability Theory and Related Fields* 78 (4), pp. 509–521.
- Bai, Z. D., Silverstein, J. W., and Yin, Y. Q. (1988). “A note on the largest eigenvalue of a large dimensional sample covariance matrix”. In: *J. Multivar. Anal.* 26.2, pp. 166–168.
- Benaych-Georges, F. and Nadakuditi, R. R. (2012). “The singular values and vectors of low rank perturbations of large rectangular random matrices”. In: *J. Multivar. Anal.* 111, pp. 120–135.

- Borell, C. (1975). “The Brunn-Minkowski inequality in Gauss space”. In: *Inventiones Mathematicae* 30 (2), pp. 207–216.
- Etemadi, N. (1981). “An elementary proof of the strong law of large numbers”. In: *Probability Theory and Related Fields* 55 (1), pp. 119–122.
- Latała, R. (2005). “Some estimates of norms of random matrices”. In: *Proc. Amer. Math. Soc.* 133.5, pp. 1273–1282.
- Rohde, A. (2012). “Accuracy of empirical projections of high-dimensional Gaussian matrices”. <http://arxiv.org/pdf/1107.5481>.
- Sudakov, V. N. and Tsirel’son, B. S. (1974). “Extremal properties of half-spaces for spherically invariant measures. (Russian)”. In: *Zap. Naunc. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* 41. (Russian), pp. 14–24.
- Thrum, R. (1987). “A remark on almost sure convergence of weighted sums”. In: *Probability Theory and Related Fields* 75 (3), pp. 425–430.